Available at WWW.MATHEMATICSWEB.ORG POWERED BY SCIENCE @ DIRECT•

ELSEVIER

Linear Algebra and its Applications 360 (2003) 1–13

LINEAR ALGEBRA AND ITS APPLICATIONS

www.elsevier.com/locate/laa

Schur majorization inequalities for symmetrized sums with applications to tensor products[☆]

James V. Bondar

School of Mathematics and Statistics, Carleton University, Ottawa, Ont., Canada K1S 5B6

Received 30 January 2001; accepted 21 May 2002

Submitted by C. Davis

Abstract

We show that if $w \prec y$ and $x \prec z$ are four vectors in R^n, then a number of Schur majorizations hold between "symmetrized" vector functions of w, x, y and z, e.g., $(w_i + x_j)_{i,j} \prec (y_i + z_j)_{i,j}$ where the left-hand expression means the vector of dimension n^2 consisting of all sums $w_i + x_j$ of the co-ordinates of w and x, arranged in lexicographic order. Among other things, we get vector and matrix versions of Muirhead's theorem for scalar inequalities. From the vector inequalities follow many scalar inequalities for "symmetrized" sums, some of which are scattered through the inequality literature.

In Section 2, applications are given to matrix inequalities for tensor products, e.g., if A, B and C are Hermitian and $\lambda(A) \prec \lambda(B)$, then $\lambda(A \otimes C) \prec \lambda(B \otimes C)$.

AMS classification: 15A45; 26XX

Keywords: Schur majorization; Eigenvalue inequalities; Tensor products

1. Scalar and vector inequalities

This section presents some new majorization inequalities for vectors of "symmetrized" sums and products. Section 2 gives applications to linear algebra of the inequalities of Section 1; these are eigenvalue and norm inequalities, mainly for tensor products. The inequalities of Section 1 also have applications in probability to the

☆ Research supported by a GR5 grant from Carleton University.
E-mail address: jbondar@math.carleton.ca (J.V. Bondar).

PII: S0024-3795(02)00461-5

ordering of random variables, and also in statistics where such symmetrized sums are called "U-statistics"; these applications will be dealt with elsewhere.

We adopt the standard notation for Schur majorization as found in [2,4,6,7]. We shall use "x in R^n" to refer to a row vector. This paper could have been written using column vectors, but since this author hates writing transpose signs everywhere, all our vectors will be row vectors. For any $x = (x_1, \dots, x_n)$ in R^n, if f is a function from R to R, then $f(x) = (f(x_1), \dots, f(x_n))$. If x and y are vectors in R^n and R^m (resp.), then we shall create new row vectors by writing round brackets with a formula inside them involving x_i and y_j, and an inequality in i and j written as a subscript to the closing bracket to indicate the range of the formula. For example, if x is in R^n, $(x_i y_j)_{i<j}$ means the vector (*not* the matrix) ordered lexicographically from largest to smallest in i and j, consisting of all products $x_i x_j$ with $i < j$; this is thus a vector living in $R^{n(n-1)/2}$. Hence, if x in R^3 is $x = (x_1, x_2, x_3)$, then $(x_i x_j)_{i<j} = (x_1x_2, x_1x_3, x_2x_3)$. Another example: if $x \in R^n$ and $y \in R^m$, the symbol $(x_i y_j)_{i,j}$ means the row vector (*not* the matrix) of all products $x_i y_j$ with $1 \leqslant i, j \leqslant n$, also ordered lexicographically into a vector in R^{nm}. Similarly for vectors involving more than two subscripts. With a little abuse of language, the vector $(x_i y_j)_{i,j}$ may be thought of as the tensor product $x \otimes y$, $(x_i x_j)_{i \leqslant j}$ as the symmetric tensor product $x \vee x$, and $(x_i x_j)_{i<j}$ as the antisymmetric tensor product $x \wedge x$.

If x and y are in R^n, we write $x \ll y$ if $x_i \leqslant y_i$ $(i = 1, \dots, n)$. A function $\Phi : R^{n_1} \times \cdots \times R^{n_p} \to R^m$ is affine in its first co-ordinate if $\Phi(\alpha x + \beta y, u, v, \dots) = \alpha\Phi(x, u, v, \dots) + \beta\Phi(y, u, v, \dots)$ (all $\alpha, \beta \geqslant 0$, $\alpha + \beta = 1$, $x, y \in R^{n_1}$, $u \in R^{n_2}$, $v \in R^{n_3}, \dots$), and convex (resp. concave) if "=" is replaced by "$\ll$" (resp. "$\gg$"). Similarly for the ith coordinate of Φ for $i > 1$. $\mathscr{S}_n$ sometimes denotes the group of permutations of the integers 1–n, and sometimes the group of permutations of the co-ordinates of R^n; the context will make clear which is meant. Φ is $\mathscr{S}_n$-convex in its 1st co-ordinate if

$$\Phi\left(\sum_i \alpha_i \pi_i x, y, \dots, z\right) \ll \sum_i \alpha_i \Phi(\pi_i x, y, \dots, z)$$
$$\left(\pi_i \in \mathscr{S}_n, \alpha_i \geqslant 0, \sum \alpha_i = 1\right);$$

Φ is $\mathscr{S}_n$-concave if "$\ll$" is replaced by "$\gg$" in the last inequality. Comment: by the theory of Marshall and Olkin [7, 2.B.1], it is sufficient to check $\mathscr{S}_n$-convexity for the case $\sum \alpha_i \pi_i x = \alpha_1 x + \alpha_2 \pi x$, where π is an arbitrary transposition. (Marshall and Olkin call such a transformation a T-transform.)

Theorem 1.1. *Let $\Phi : R^{n_1} \times \cdots \times R^{n_p} \to R^m$ be a function of p vector arguments which is linear or affine in its first argument, and is such that, for any π in $\mathscr{S}_{n_1}$, and x in R^{n_1}, y in $R^{n_2}, \dots, z$ in R^{n_p}, there is a Π in $\mathscr{S}_m$ such that*

$$\Phi(\pi x, y, \dots, z) = \Pi\Phi(x, y, \dots, z), \tag{1.1}$$

and let $x^{(1)}$ and $x^{(2)}$ be two vectors in R^n such that

$$x^{(1)} \prec x^{(2)}. \tag{1.2}$$

Then we have

$$\Phi(x^{(1)}, y, \ldots, z) \prec \Phi(x^{(2)}, y, \ldots, z). \tag{1.3}$$

If Φ is $\mathscr{S}_n$-convex in its 1st *co-ordinate* (*instead of linear or affine*), *then* "$\prec$" *must be replaced by* "$\prec_w$" *in inequality* (1.3); *if Φ is $\mathscr{S}_n$-concave, then* "$\prec$" *must be replaced by* "$\prec^w$". *Similar statements hold of course for the other* (*vector-valued*) *co-ordinates $y, \ldots, z$.*

Proof. If $p = 1$, Theorem 1.1 is a theorem of [2, Theorem 2.1], see also [4, Theorem II.3.3]. The proof of Theorem 1.1 follows from fixing $y, \ldots, z$ and applying Ando's theorem to Φ and x. $\square$

Numerous inequalities follow from the theorem; for simplicity we shall state most of these for the case of $p = 2$, that is, for expressions with only two subscripts, however, they are valid for arbitrary positive integers p. Let w and x be in R^m and R^n, respectively, and $\Phi(w, x)$ be the vector $(w_i + x_j)_{i,j}$ in R^{mn}; since the map $(w, x) \to w + x$ is affine in each of w and x, Theorem 1.1 then yields (1.5a); its multiplicative analogue (1.5b) follows from the fact that

$$(w, x) \to (w_i x_j)_{i,j} = (w_1x_1, \ldots, w_1x_n, w_2x_1, \ldots, w_mx_n) \tag{1.4}$$

is linear (hence affine) in each of x and y.

Corollary 1.2. *Let $w \prec y$ be two vectors in R^m and $x \prec z$ be two vectors in R^n. Then*

$$(w_i + x_j)_{i,j} \prec (y_i + z_j)_{i,j}, \tag{1.5a}$$

$$(w_i x_j)_{i,j} \prec (y_i z_j)_{i,j}, \tag{1.5b}$$

$$(w_i - x_j)_{i,j} \prec (y_i - z_j)_{i,j}, \tag{1.5c}$$

$$(w_i / x_j)_{i,j} \prec_w (y_i / z_j)_{i,j} \tag{1.5d}$$

(*provided all w_i, x_j, y_i and z_j in* (1.5d) *are* > 0).

We remind the reader of our convention that the expression on each side of the inequalities (1.5a)–(1.5d) is a row vector in R^{mn} containing the quantities of the indicated form, arranged in lexicographic order of i and j. Thus in (1.5a), on the left-hand side we have

$$(w_i + x_j)_{i,j} = (w_1 + x_1, \ldots, w_1 + x_n, w_2 + x_1, \ldots, w_2 + x_n, \ldots, w_m + x_n).$$

Corollary 1.3. *Let $w \prec y$ be vectors in R^n. Then*

$$(w_i + w_j)_{i<j} \prec (y_i + y_j)_{i<j} \tag{1.6a}$$

$$(w_{i_1} w_{i_2} \cdots w_{i_p})_{i_1 < i_2 < \cdots < i_p} \prec^w (y_{i_1} y_{i_2} \cdots y_{i_p})_{i_1 < i_2 \cdots < i_p} \tag{1.6b}$$

for $n \geqslant p \geqslant 2$ if all w_i and y_i are non-negative; *if $p = 2$ this holds without the non-negativity assumption*

$$(|w_i - w_j|)_{i<j} \prec_w (|y_i - y_j|)_{i<j}, \tag{1.6c}$$

$$(w_i / w_j)_{i<j} \prec_w (y_i / y_j)_{i<j} \tag{1.6d}$$

if w and y are in $\mathscr{D}_{++}$ (i.e., co-ordinates in descending order and strictly positive).

Note that (1.6b) requires different assumptions for the case $p = 2$ and for the case $p > 2$ ((1.6b) is the only inequality in Corollaries 1.2–1.4 which requires different assumptions for different values of p). We shall postpone the proof of inequality (1.6b) until a little later; it follows from Proposition 1.5(B) with $l = p$ and $f(x) = x_1 x_2 \cdots x_p$, which is not necessarily Schur-concave if $p \geqslant 3$ and w and y have negative co-ordinates, but is always Schur-concave for $p = 2$. Note the partial lack of symmetry in inequality (1.6d); it contains only those terms in $(w_i/w_j)_{i \neq j}$ which are greater than 1. No majorization exists in general for the vector $(w_i/w_j)_{i>j}$. The proof of (1.6d) will also be postponed; it follows from Proposition 1.5(A) with $l = 2$ and $f(x) = x_{[1]}/x_{[2]}$ on the set $\mathscr{A} = R^n$ ($x_{[i]}$ is the ith largest co-ordinate of x).

Corollary 1.4. *Let $w \prec y$ be two vectors in R^n. Then we have the following majorizations similar to those in Corollary 1.3, but allowing repeated subscripts*:

$$(w_i + w_j)_{i \leqslant j} \prec (y_i + y_j)_{i \leqslant j}, \tag{1.7a}$$

$$(w_i w_j)_{i \leqslant j} \prec_w (y_i y_j)_{i \leqslant j} \tag{1.7b}$$

if the co-ordinates of w and y are non-negative.

Statement (1.7a) follows directly from Theorem 1.1. Inequality (1.7b), on the other hand, is not so straightforward; a proof is given after Proposition 1.5.

Note. Replacing "$w \prec y$" by "$w \prec_w y$" (resp. "$w \prec^w y$") in the assumptions of Corollary 1.2, 1.3 or 1.4 changes the "$\prec$" in inequality "a" of that corollary to "$\prec_w$" (resp. "$\prec^w$"). If the co-ordinates of x and z are all > 0, then changing the "$\prec$" in the assumptions of Corollary 1.2 to "$\prec_w$" (resp. "$\prec^w$") changes the majorization in inequality (1.5b) in the same way. If the co-ordinates of w and y are all > 0, then changing the "$\prec$" in the assumptions of Corollary 1.3 to "$\prec^w$" leaves the "$\prec^w$" unchanged in inequality (1.6b). If the co-ordinates of w and y are all > 0, then changing the "$\prec$" in the assumptions of Corollary 1.4 to "$\prec_w$" leaves the "$\prec_w$" unchanged in inequality (1.7b).

Formula (1.6b) may be viewed as a strengthening of the fact that the symmetric function is Schur concave, while (1.7b) strengthens the fact that the complete symmetric function $C_k = \sum x_1^{i_1} \cdots x_n^{i_n}$ (sum over $i_1 + \cdots + i_n = k,\ i_j \geqslant 0$) is Schur-convex.

Proposition 1.5. (A) *Let $\mathscr{A}$ be a symmetric convex set in R^l and f be a Schur-convex function defined on $\mathscr{A}$ with the property that for each fixed $(x_2, \dots, x_l)$, the function $f(z, x_2, \dots, x_l)$ is convex in z in the set $\{z \mid (z, x_2, \dots, x_l) \in \mathscr{A}\}$. For any $n > l$, define the vector*

$$F(x_1, \dots, x_n) = (f(x_{\pi(1)}, \dots, x_{\pi(l)}))_{\pi \in \mathscr{S}_n}, \tag{1.8a}$$

then $x, y \in \mathscr{A}$ such that $x \prec y$ implies that

$$F(x) \prec_w F(y). \tag{1.8b}$$

Consequently, the real-valued sum

$$\sum_{\pi \in \mathscr{S}_n} f(x_{\pi(1)}, \dots, x_{\pi(l)}) \quad \text{is Schur-convex.} \tag{1.8c}$$

(B) *If the Schur-convexity and convexity of part* (A) *are changed to Schur-concavity and concavity, then the conclusion* (1.8b) *is changed to $F(x) \prec^w F(y)$, and* (1.8c) *is Schur-concave.*

(C) *In the hypotheses of* (A) *of the theorem, if "$x \prec y$" is weakened to "$x \prec_w y$", and we add the condition "f is a monotone increasing function", then the conclusion* (1.8b) *still holds.*

Note. In [7, Proposition 3.G.3], Proposition 1.5 is given in its scalar form (1.8c), not the vector form (1.8b).

If $G(x)$ is defined as $(f(z))_{z \in C(x,n,l)}$ where z runs through the set $C(x, n, l)$ of all the n choose l subsets of l elements taken from the co-ordinates of x (we can think of the subsets z as ordered lexicographically), then (1.8b) is equivalent to $G(x) \prec_w G(y)$, and (1.8c) to $\sum_{C(x,n,l)} f(x) \leqslant \sum_{C(y,n,l)} f(y)$. Note that $F(x)$ is equal (up to permutation of co-ordinates) to the direct sum of $l!$ copies of $G(x)$, but nevertheless, $F(x) \prec_w F(y)$ is equivalent to $G(x) \prec_w G(y)$, and expression (1.8c) is equal to $(l!) \sum_{C(x,n,l)} f(x)$. (The notation $C(x, n, l)$ is essentially introduced in [7, Chapter 3.G].)

Proof. (A) for $l = 2$ (the proof is similar for $l > 2$). Let x and y be in R^n with $x \prec y$. By the remark just before Theorem 1.1, there is a sequence of vectors $x^{(i)}$ in R^n, of π_i in $\mathscr{S}_n$, α_i and β_i, such that $x = x^{(N)} \prec x^{(N-1)} \prec \cdots \prec x^{(1)} \prec x^{(0)} = y$ and

$$x^{(i+1)} = \alpha_i x^{(i)} + \beta_i \pi_i x^{(i)}, \quad i = 0, \dots, N-1, \tag{1.8d}$$

where $\alpha_i + \beta_i = 1$, $\alpha_i, \beta_i \geqslant 0$ and each π_i is a transposition. If we can prove that (1.8d) implies $F(x^{(i+1)}) \prec^w F(x^{(i)})$, then Proposition 1.5 follows. Hence, it suffices to prove that if $x, y \in R^n$ where x equals the convex linear combination $\alpha y + \beta \pi y$ and π is a transposition, then there exists u such that $F(x) \ll u \prec F(y)$. For simplicity we consider the case in which π is the transposition interchanging x_1 and x_2, but the proof holds for all transpositions π. We now have

$$F(x) = \big(f(x_1, x_2), f(x_1, x_3), \ldots, f(x_{n-1}, x_n)\big),$$

where

$$\begin{aligned} f(x_1, x_2) &= f(\alpha y_1 + \beta y_2, \alpha y_2 + \beta y_1) \\ &= f(\alpha(y_1, y_2) + \beta(y_2, y_1)) \leqslant f(y_1, y_2) \end{aligned}$$

by the Schur-convexity of f. For $j \geqslant 3$, we have

$$f(x_1, x_j) = f(\alpha y_1 + \beta y_2, y_j) \leqslant \alpha f(y_1, y_j) + \beta f(y_2, y_j)$$

and

$$f(x_2, x_j) = f(\alpha y_2 + \beta y_1, y_j) \leqslant \alpha f(y_2, y_j) + \beta f(y_1, y_j)$$

by the convexity of f in each co-ordinate. For $i, j \geqslant 3$, we have $f(x_i, x_j) = f(y_i, y_j)$. Hence, the vector $F(x)$ is

$$\begin{aligned} &\ll (\alpha f(y_1, y_2) + \beta f(y_1, y_2), \\ &\quad \alpha f(y_1, y_3) + \beta f(y_2, y_3), \ldots, \alpha f(y_1, y_n) + \beta f(y_2, y_n), \\ &\quad \alpha f(y_2, y_3) + \beta f(y_1, y_3), \ldots, \alpha f(y_2, y_n) + \beta f(y_1, y_n), \\ &\quad \alpha f(y_3, y_4) + \beta f(y_3, y_4), \ldots, \alpha f(y_{n-1}, y_n) + \beta f(y_{n-1}, y_n)) \\ &= \alpha F(y) + \beta \Pi F(y) \prec F(y), \quad \text{where } \Pi \text{ is in } \mathscr{S}_{n(n-1)/2}. \end{aligned}$$

(B) Apply (A) to $-f$ and use the fact that $w \prec_w z$ implies $-w \prec^w -z$. $\square$

In preparation for the proof of Corollary 1.4(1.7b) we need to prove (1.7b) for $n = 2$.

Lemma 1.6. *If $x \prec_w y$ are two vectors in R^2 with non-negative elements, and H is the vector-valued function defined by $H(x_1, x_2) = (x_1^q, x_1^{q-1}x_2, \ldots, x_1x_2^{q-1}, x_2^q)$, then*

$$H(x) \prec_w H(y). \tag{1.9a}$$

Let $L_k(x)$ be the sum of the k largest co-ordinates of $H(x)$, $k = 1, 2, \ldots, q + 1$. If $x_1 \geqslant x_2 \geqslant 0$, then

$$\partial L_k(x)/\partial x_1 \geqslant \partial L_k(x)/\partial x_2 \geqslant 0 \quad \textit{for all } k. \tag{1.9b}$$

Proof. If we can show that $L_k(x)$ is Schur-increasing in x for all k, then $L_k(x) \leqslant L_k(y)$, so by definition, $H(x) \prec H(y)$. For simplicity, we assume that $x_1 \geqslant x_2$, so that $H(x)$ is in decreasing order, and $L_k(x) = \sum_{i=0}^{k-1} x_1^{q-i} x_2^i$. By Schur's criterion [7, 3.4.A], it will be sufficient to show that $D_1L_k(x) \geqslant D_2L_k(x) \geqslant 0$, where $D_s = \partial/\partial x_s$, $s = 1, 2$.

Now,

$$D_1L_k(x) = qx_1^{q-1} + (q-1)x_1^{q-2}x_2 + \cdots + (q-k+1)x_1^{q-k}x_2^{k-1}$$

while

$$D_2L_k(x) = x_1^{q-1} + \cdots + (k-1)x_1^{q-k+1}x_2^{k-2}.$$

This expression for $D_2L_k(x)$ is the inner product $\langle(1, 2, \ldots, k-1, 0) \mid u\rangle$, where $u = (x_1^{q-1}, \ldots, x_1^{q-k}x_2^{k-1})$. Then

$$\begin{aligned}\langle(1, 2, \ldots, k-1, 0) \mid u\rangle &\leqslant \langle(q-k+1, q-k+2, \ldots, q) \mid u\rangle \\ &\leqslant \langle(q, \ldots, q-k+2, q-k+1) \mid u\rangle\end{aligned}$$

(since the inner product is larger when the vectors are similarly ordered [7, 6.A.3]). Since this last inner product is equal to $D_1L_k(x)$, both (1.9a) and (1.9b) are proved. □

Proof of inequality (1.7b) (*cf. Corollary* 1.4). Let $x \prec y$ be in R^n, and for each $k = 1, 2, \ldots, \dim(G)$, let $M_k(x)$ be the sum of the k largest co-ordinates of $G(x) = (x_{i_1}x_{i_2}\cdots x_{i_p})_{i_1\leqslant i_2\cdots\leqslant i_p}$. As with Lemma 1.6, if we can show that $M_k(x)$ is Schur-increasing in x for all k, then $M_k(x) \leqslant M_k(y)$, so by definition $G(x) \prec_w G(y)$. For simplicity, we shall assume that the entries in the vectors x and y have been arranged in decreasing order.

Fix s such that $1 \leqslant s \leqslant n-1$. If C is a product of factors x_i with i not equal to s or $s+1$, we say that C is of zero degree. For any k, if $M_k(x)$ contains the term $Cx_s^r x_{s+1}^{q-r}$ for C of zero degree, then since $x_s \geqslant x_{s+1}$, $M_k(x)$ must also contain $Cx_s^i x_{s+1}^{q-i}$ for all $i > r$. Thus, $M_k(x)$ is the sum of expressions of the form

$$E = C\left(x_s^q + x_s^{q-1}x_{s+1} + \cdots + x_s^{q-r}x_{s+1}^r\right)$$

for some C of zero degree and some q and r such that $1 \leqslant r \leqslant q \leqslant p$. This expression is equal to $L_{r+1}(x)$ from Lemma 1.6 (except for a change of subscripts), and by (1.9b), for each such E we have $\partial E/\partial x_s \geqslant \partial E/\partial x_{s+1} \geqslant 0$. Hence $\partial M_k(x)/\partial x_s \geqslant \partial M_k(x)/\partial x_{s+1} \geqslant 0$ for all k and for $s = 1, \ldots, n-1$; therefore $M_k(x)$ is Schur-convex by Schur's criterion [7, 3.A.4] hence $G(x) \prec_w G(y)$. □

Corollary 1.7. *If $x \prec y$ are in R^n and ψ is a convex function from R^h to R, then*

$$\left(\psi(x_{\pi(1)}, \ldots, x_{\pi(h)})\right)_{\pi\in\mathscr{S}_n} \prec_w \left(\psi(y_{\pi(1)}, \ldots, y_{\pi(h)})\right)_{\pi\in\mathscr{S}_n}. \tag{1.10a}$$

If ψ is increasing and convex, then "$x \prec y$" must be replaced by "$x \prec_w y$"; if ψ is concave, then the "$\prec_w$" in (1.10a) *must be replaced by "$\prec^w$".*

Corollary 1.8. *For x in R^n and $1 \leqslant p \leqslant n$, define*

$$\begin{aligned}F(x) &= \left(1/x_{i_1} + 1/x_{i_2} + \cdots + 1/x_{i_p}\right)_{i_1<i_2\cdots<i_p} \\ &= \left(S_{p-1}(x_{i_1}, \ldots, x_{i_p})/S_p(x_{i_1}, \ldots, x_{i_p})\right)_{i_1<i_2\cdots<i_p},\end{aligned} \tag{1.11a}$$

where $S_k(x)$ is the kth elementary symmetric function of x. F is clearly a sum of convex functions, hence Proposition 1.4 *applies and we get that expression* (1.11a) *increases in the $\prec_w$ order as x increases in the Schur majorization order. Also, the sum of the elements of the vector* (1.11a) *is Schur-convex.*

Compare Corollary 1.8 to [7, 3.G.3.b] where the corresponding result is proved for $\sum(S_1(x_{i_1},\ldots,x_{i_l})/S_l(x_{i_1},\ldots,x_{i_l}))$, where the sum is over $i_1 < i_2 < \cdots < i_l$ (equivalently, sum over $C(x,n,l)$). It may be that a similar result can be proved for $\sum(S_k(x_{i_1},\ldots,x_{i_l})/S_p(x_{i_1},\ldots,x_{i_l}))$ for all $1 \leqslant k < p \leqslant l \leqslant n$.

Now we can prove most of the inequalities in [7, Chapter 3.G], and derive vector majorization results along the way. That is, where Marshall and Olkin prove an inequality of the form $\sum_S f(x_{i_1},\ldots,x_{i_p}) \leqslant \sum_S f(y_{i_1},\ldots,y_{i_p})$ for some set S of indices, we are able to prove a vector majorization of the form $(f(x_{i_1},\ldots,x_{i_p}))_S \prec_w (f(y_{i_1},\ldots,y_{i_p}))_S$ which implies the inequality. For example:

If $\alpha, x \in R^n$ and we apply Theorem 1.1 to the mapping

$$(\alpha, x) \to \left(\sum_{i=1}^{n} \alpha_{\pi(i)} x_i\right)_{\pi\in\mathscr{S}_n},$$

then we find that $\alpha \prec \beta$ implies that

$$\left(\alpha_{\pi(1)}x_1 + \cdots + \alpha_{\pi(n)}x_n\right)_{\pi\in\mathscr{S}_n} \prec \left(\beta_{\pi(1)}x_1 + \cdots + \beta_{\pi(n)}x_n\right)_{\pi\in\mathscr{S}_n}. \tag{1.12a}$$

If we apply (1.12a) with $x = \ln y$ and exponentiate both sides of this inequality, we get:

Corollary 1.9 (Vector version of Muirhead's theorem). *If $\alpha \prec \beta$ and if $y \in R^n$ is a vector of strictly positive numbers, then*

$$\left(y_1^{\alpha_{\pi(1)}} y_2^{\alpha_{\pi(2)}} \cdots y_n^{\alpha_{\pi(n)}}\right)_{\pi\in\mathscr{S}_n} \prec_w \left(y_1^{\beta_{\pi(1)}} y_2^{\beta_{\pi(2)}} \cdots y_n^{\beta_{\pi(n)}}\right)_{\pi\in\mathscr{S}_n}. \tag{1.12b}$$

We also have

$$\left(y_{i_1}^{\alpha_{\pi(1)}} \cdots y_{i_n}^{\alpha_{\pi(n)}}\right)_{\pi\in\mathscr{S}_n} \prec \left(y_{i_1}^{\beta_{\pi(1)}} \cdots y_{i_n}^{\beta_{\pi(n)}}\right)_{\pi\in\mathscr{S}_n} \tag{1.12c}$$

which is a vector Muirhead's theorem with repetition of indices.

The proof uses the following corollary.

Corollary 1.10. *For $x \in R^n$ and $a \in R^p$, define*

$$F(a) = (a_1 x_{i_1} + a_2 x_{i_2} + \cdots + a_p x_{i_p})_{i_j=1}^{n}$$

in R^{n^p} where each symbol x_{i_j} runs through each one of the n values $x_1,\ldots,x_n$. If $a \prec b$, then $F(a) \prec F(b)$.

This is a direct application of Theorem 1.1 to $F(a)$. It has implications in statistics: if $X_1,\ldots,X_n$ are independent identically distributed symmetric random variables, then the distribution of $\sum_i a_i X_i$ becomes more "peaked" (less disperse) as the vector a increases in the Schur majorization order.

Corollary 1.11 (Rousseau). *If $\varphi : R \to R$ is convex, then*

$$F(x, y) = \sum_{i,j} \varphi(x_i - y_j) \quad \text{and} \quad G(x, y) = \sum_{i<j} \varphi(x_i - x_j)$$

are Schur-convex functions of x and of y (x and y being vectors in R^n and R^m, respectively).

This corollary was originally proved by Ronald Rousseau in connection with his work on measures of diversity and of concentration. See [5], especially their Corollary 10, for a discussion. While Corollary 1.11 may be deduced from Theorem 3.G.1.b of [7] with some effort, it deserves to be stated as a result on its own. It follows from our Corollary 1.2 and the fact that a convex function f of a single real variable maps vectors $w \prec y$ into vectors $f(w) = (f(w_1), \ldots, f(w_n)) \prec_w f(y) = (f(y_1), \ldots, f(y_n))$.

Corollary 1.12. *If $w \prec y$ and $x \prec z$ are four vectors in R^n, then*

$$\left(w_i^p x_j^q\right)_{i,j} \prec_w \left(y_i^p z_j^q\right)_{i,j} \quad \text{if } 1 < p, q < \infty. \tag{1.13a}$$

If the condition "$1 < p, q < \infty$" is replaced by "$0 < p, q < 1$", then the inequality "$\prec_w$" must be replaced by "$\prec^w$".

$$\left(w_i^p x_j^q\right)_{i,j} \prec \left(y_i^p z_j^q\right)_{i,j} \quad \text{if } p, q \in A \text{ and all } w_i, x_j, y_i, z_j > 0, \tag{1.13b}$$

where A is the union of the intervals $(-\infty, 0]$ and $[1, \infty)$.

Log-majorization results follow from the corollaries above.

Corollary 1.13. *Let w and y be in R^n and x and z be in R^m. If $\ln w \prec \ln y$ and $\ln x \prec \ln z$, (which imply that the entries of w, x, y and z are > 0), then Corollaries 1.2 and 1.3 imply:*

$$\ln(w_i x_j)_{i,j} \prec \ln(y_i z_j)_{i,j} \tag{1.14a}$$

from which follows $(w_i x_j)_{i,j} \prec_w (y_i z_j)_{i,j}$ and also

$$\left(\sum w_i\right)\left(\sum x_j\right) \leqslant \left(\sum y_i\right)\left(\sum z_j\right);$$

$$\ln(w_i/x_j)_{i,j} \prec \ln(y_i/z_j)_{i,j}, \tag{1.14b}$$

$$\ln(w_i w_j)_{i<j} \prec \ln(y_i y_j)_{i<j} \quad \text{and} \tag{1.14c}$$

$$(w_i w_j)_{i<j} \prec_w (y_i y_j)_{i<j}. \tag{1.14d}$$

From (1.14d), *it follows that* $\ln w \prec \ln y$ *implies* $S_k(w) \leqslant S_k(y)$, $k = 1, \dots, n$, *where* S_k *is the* k*th elementary symmetric function. This contrasts with the fact that* $w \prec y$ *implies* $S_k(w) \geqslant S_k(y)$.

Corollary 1.14. *If* $\ln x \prec \ln z$ *and* $a \prec b$ *are four vectors in* R^n (*with all* $x_i, z_j > 0$), *then*

$$(a_i \ln x_j)_{i,j} \prec (b_i \ln z_j)_{i,j} \tag{1.15a}$$

which in turn implies

$$(x_j^{a_i})_{i,j} \prec^w (z_j^{b_i})_{i,j}. \tag{1.15b}$$

2. Applications to matrix inequalities

Several results on tensor products follow immediately from the corollaries of Section 1. For any matrix A, $s(A)$ denotes the vector of singular values of A, and $\lambda(A)$ denotes the vector of eigenvalues; both vectors ordered from largest to smallest. $\vee$ and $\wedge$ are the symmetric and antisymmetric (exterior) tensor products, respectively ($\wedge^k A$ is also known as the kth compound of A or $A^{(k)}$); these products have eigenvalues which are symmetrized sums and products of the types discussed in Section 1. Directly from (1.5b), (1.6b) and (1.7b) we get Theorems 2.1 and 2.2.

Theorem 2.1. *If* $w \prec y$ *and* $x \prec z$ *are four vectors in* R^n, *then we have* $w \otimes x \prec y \otimes z$. *If the elements of* w, x, y *and* z *are also nonnegative*, *we have* $w \wedge x \prec^w y \wedge z$ *and* $w \vee x \prec_w y \vee z$.

Theorem 2.2. *Let* A, B *and* C *be matrices and let* $s(A) \prec s(B)$. *Then*

$$s(A \otimes C) \prec s(B \otimes C), \tag{2.1a}$$

$$s(\wedge^k A) \prec^w s(\wedge^k B), \quad k = 1, 2, \dots, \quad \textit{and} \tag{2.1b}$$

$$s(\vee^k A) \prec_w s(\vee^k B), \quad k = 1, 2, \dots \tag{2.1c}$$

If A, B *and* C *are Hermitian matrices such that* $\lambda(A) \prec \lambda(B)$, *then*

$$\lambda(A \otimes C) \prec \lambda(B \otimes C), \tag{2.1d}$$

$$\lambda(\wedge^k A) \prec^w \lambda(\wedge^k B) \tag{2.1e}$$

holds for $k = 2$; *it holds for* $k = 3, 4, \dots,$ *if* A *and* B *are positive semidefinite*

$$\lambda(\vee^k A) \prec_w \lambda(\vee^k B), \quad k = 1, 2, \dots, \tag{2.1f}$$

if A *and* B *are positive semidefinite*.

If $p(s,t)=\sum_{k,l} d_{k,l}s^k t^l$ *is a polynomial in s and t, and* $P(A,B)=\sum_{k,l} d_{k,l}A^k \otimes B^l$ *with* $\alpha=\lambda(A) \prec \beta=\lambda(B)$, *then* $\lambda P(A,B)=(p(\alpha_i,\beta_j))_{i,j}$; *and by Theorem* 1.1,

$$\lambda P(A,C) \prec_w \lambda P(B,C) \tag{2.1g}$$

under the assumptions of (2.1d), *if* $p(\cdot,\cdot)$ *is convex in the convex hull of the* $(\beta_i\gamma_j, \beta_k\gamma_l)$ $(\gamma=\lambda(C))$.

For any pair of matrices C and D, $s(C) \prec_w s(D)$ is equivalent to $|||C||| \leqslant |||D|||$ for every unitarily invariant norm $|||\cdots|||$, by the Ky Fan dominance theorem [2, IV 2.2]. Thus the conditions of Theorem 2.2 imply that $|||A\otimes C||| \leqslant |||B \otimes C|||$ and $|||\vee^k A||| \leqslant |||\vee^k B|||$ for all unitarily invariant norms; similar implications follow for many of the following corollaries.

Corollary 2.3. *Let the matrix B be partitioned as*

$$\begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}.$$

Now set

$$A=\begin{pmatrix} B_{11} & O \\ O & B_{22} \end{pmatrix}$$

(*a pinching of B in the sense of Bhatia* [2]). *Then* (2.1a)–(2.1c) *hold if B and C are arbitrary square matrices*, (2.1d) *holds for Hermitian matrices, and* (2.1e) *and* (2.1f) *hold for positive semidefinite matrices. In particular*, (2.1e) *implies that* $\mathrm{tr}(\wedge^k A) \geqslant \mathrm{tr}(\wedge^k B)$, *which is already known* [2, II.43], *and* $\mathrm{tr}(\wedge^k A^{-1}) \leqslant \mathrm{tr}(\wedge^k B^{-1})$.

Proof. By assumption, $s(A) \prec_w s(B)$ [2, II.38]; now apply Theorem 2.2. □

From the vector Muirhead's theorem (Corollary 1.9), we get:

Corollary 2.4 (Matrix version of Muirhead's theorem). *Let* $x \prec y$ *be vectors in* R^n, *and A be an* $n \times n$ *positive definite matrix. Then*

$$\lambda(A^{x_1}\otimes A^{x_2}\otimes\cdots\otimes A^{x_n}) \prec \lambda(A^{y_1}\otimes\cdots\otimes A^{y_n}).$$

Corollary 2.5. *If* A, B *and* C *are* $n\times n$ *matrices with n real eigenvalues, for which* $\lambda(A)\prec\lambda(B)$, *then*

$$\lambda(A\otimes I+I\otimes C) \prec \lambda(B\otimes I+I\otimes C), \quad \textit{and} \tag{2.4a}$$

$$\lambda(A^{[k]}) \prec \lambda(B^{[k]}) \tag{2.4b}$$

for $k=1,2,\ldots,n$, *where* $A^{[k]}$ *is the kth additive compound of A.*

Theorem 2.6. *Let* A, B, C, D *be positive semidefinite matrices such that* $\lambda(A)\prec\lambda(B)$ *and* $\lambda(C)\prec\lambda(D)$. *Then*

$$\lambda(A^p \otimes C^q) \prec^w \lambda(B^p \otimes D^q) \quad \text{if } 0 \leqslant p, q \leqslant 1, \tag{2.5a}$$

$$\lambda(A^p \otimes C^q) \prec_w \lambda(B^p \otimes D^q) \quad \text{if } 1 \leqslant p, q < \infty. \tag{2.5b}$$

If we strengthen the above conditions to make A, B, C and D positive definite, then we have

$$\lambda(A^{-p} \otimes C^{1+q}) \prec_w \lambda(B^{-p} \otimes D^{1+q}) \quad \text{if } 0 \leqslant p, q < \infty. \tag{2.5c}$$

Proof. Under the assumptions of the theorem, if $0 \leqslant p, q \leqslant 1$, then $\lambda(A^p) \prec^w \lambda(B^p)$ and $\lambda(C^q) \prec^w \lambda(D^q)$. By (1.5b), we get (2.5a). Similarly for (2.5b) and (2.5c). □

The tensor products in Theorem 2.6 are known to have convexity or concavity properies in the Loewner (positive definite) ordering " $<_{\mathrm{L}}$" ([1] or [3, Section 3]). This is not mere co-incidence: Loewner-convexity in an argument of an expression like those in Theorem 2.6 implies a variant of weak Schur convexity. To be precise:

Theorem 2.7. *Let the map $A \to F(A)$ be defined for positive semidefinite A and take Hermitian values. Let $F(A)$ be convex in A in the Loewner order, i.e. $F(\alpha A + \gamma C) <_{\mathrm{L}} \alpha F(A) + \gamma F(C)$, $(\alpha, \gamma \geqslant 0$ and $\alpha + \gamma = 1)$. Further, let $\lambda F(A) = \lambda F(UAU^{-1})$ for all unitary U and positive semidefinite A. Then $\lambda(A) \prec \lambda(B)$ implies that $\lambda F(A) \ll \lambda F(B)$, hence $\lambda F(A) \prec_w \lambda F(B)$. If F is concave, then "$\ll$" and "$\prec_w$" are replaced by "$\gg$" and "$\prec^w$" respectively.*

Proof. Let $\lambda(A) \prec \lambda(B)$, and let F obey the conditions in Theorem 2.7. Then there exist α_i and unitary matrices U_i such that A is equal to the convex linear combination $\sum \alpha_i U_i B U_i^{-1}$ (see, for example, [8, p. 136]). Now by convexity, $F(A) = F(\sum \alpha_i U_i B U_i^{-1}) <_{\mathrm{L}} \sum \alpha_i F(B)$. Therefore, $\lambda F(A) \ll \lambda(\sum \alpha_i F(B)) = \lambda F(B)$. □

Ando's theorems prove joint convexity and concavity (which is stronger than our separate convexity and concavity) in the Loewner order (which is stronger than our "$\prec_w$" order). This is why the conditions of our Theorem 2.6 are weaker than Ando's (for example, each side of formula (2.5a) requires the condition $p + q = 1$ for concavity, but $0 \leqslant p, q \leqslant 1$ is all that is is needed for our inequality (2.5a)). By Theorem 2.7, Theorem 2.6 can be strengthened with "$\ll$" replacing "$\prec_w$" and "$\gg$" replacing "$\prec^w$".

References

[1] T. Ando, Concavity of certain maps on positive definite matrices and applications to Hadamard products, Linear Algebra Appl. 26 (1979) 203–241.

[2] T. Ando, Majorization, doubly stochastic matrices and comparison of eigenvalues, Linear Algebra Appl. 118 (1989) 163–248.

[3] T. Ando, Operator-Theoretic Methods for Matrix Inequalities, Hokusei Gakuen University, 1998.
[4] R. Bhatia, Matrix Analysis, Springer, Berlin, 1997.
[5] L. Egghe, R. Rousseau, Transfer principles and a classification of concentration measures, J. Am. Soc. Inform. Sci. 42 (1991) 479–489.
[6] R.A. Horn, C.R. Johnson, Topics in Matrix Analysis, Cambridge, 1991.
[7] A. Marshall, I. Olkin, Inequalities: Theory of Majorization and its Applications, Academic Press, New York, 1979.
[8] H. Miranda, R.C. Thompson, Group majorization, the convex hulls of sets of matrices and the diagonal element-singular value inequalities, Linear Algebra Appl. 199 (1994) 131–141.